

AD-A078 490

MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 12/1
MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING.(U)

SEP 79 S K MITTER , D OONE

AFOSR-77-3281

JNCLASSIFIED

LIDS-P-943

AFOSR-TR-79-1256

NL

1 of 1

AD
AO 78490



END
DATE
FILMED

1-80

DDC

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

LEVEL II

| | | | | | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|--|-------------------------------------------------------------------------------|--|--------------------------------------------|--|
| 18. REPORT DOCUMENTATION PAGE | | 2. GOVT ACCESSION NO. | | 3. RECIPIENT'S CATALOG NUMBER | |
| 1. REPORT NUMBER AFOSR-TR-79-1256 | | | | | |
| 4. TITLE (and Subtitle) MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING | | 5. TYPE OF REPORT & PERIOD COVERED Interim rept. | | 6. PERFORMING ORG. REPORT NUMBER | |
| 7. AUTHOR(s) S.K. Mitter D./Ocone | | 8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3281 | | | |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Laboratory for Information & Decision Systems Cambridge, MA 02139 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A1 | | | |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332 | | 12. REPORT DATE September 1979 | | 13. NUMBER OF PAGES 6 | |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) LIDS-P-943 | | 15. SECURITY CLASS. (of this report) UNCLASSIFIED | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE | |

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
PRESENTED AT IEEE CONFERENCE ON DECISION AND CONTROL

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
In their seminal paper, Fujisaki, Kallianpur and Kunita [1] showed how the best least squares estimate of a signal contained in additive white noise can be represented as a stochastic integral with respect to innovation process, the integral being adapted to the observation process. The difficulty with this representation is that in general this estimate is not useful for computing the estimate since the innovations process depends on the estimate of the signal itself. In this paper we discuss representation of the estimate directly in terms of the observation process. In doing so, we derive new results on

DD FORM 1 JAN 73 1473

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 950

ADA 078490

DDC FILE COPY

DDC
RECEIVED
DEC 20 1979
B

20. Abstract continued.

multiple integral expansions for square-integrable functionals of the observation process and show the connection of this work to the theory of contraction operators on Fock space. This letter development is due to Nelson and Segal.

We also present several applications of these results to determining sub-optimal filters.

| | |
|---------------------------------|---------------------------------------------------|
| ACCESSION for | |
| NTIS | White Section <input checked="" type="checkbox"/> |
| DDC | Buff Section <input type="checkbox"/> |
| UNANNOUNCED | <input type="checkbox"/> |
| JUSTIFICATION | |
| BY | |
| DISTRIBUTION/AVAILABILITY CODES | |
| Dist. AVAIL. and/or SPECIAL | |
| A | |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING
S. K. Mitter and D. Ocone
Massachusetts Institute of Technology
Cambridge, MA

1. Introduction

In their seminal paper, Fujisaki, Kallianpur and Kunita [1] showed how the best least squares estimate of a signal contained in additive white noise can be represented as a stochastic integral with respect to innovation process, the integral being adapted to the observation process. The difficulty with this representation is that in general this estimate is not useful for computing the estimate since the innovations process depends on the estimate of the signal itself. In this paper we discuss representation of the estimate directly in terms of the observation process. In doing so, we derive new results on multiple integral expansions for square-integrable functionals of the observation process and show the connection of this work to the theory of contraction operators on Fock space. This latter development is due to Nelson and Segal.

We also present several applications of these results to determining sub-optimal filters.

2. Multiple Integrals and Filtering

In this section, we shall discuss applications of multiple integral expansions to the general filtering problem. We will consider the 'canonical' scalar filtering model:

$$y_t = \int_0^t h(x_s) ds + w_t \quad (1)$$

under the assumptions

a) x_t and w_t are independent processes

b) for some $T > 0$ $E \int_0^T h^2(x_s) ds < \infty$ (2)

c) w_t is a standard Brownian motion

If $f_t(x_{\cdot}) = f_t(x_s, s \leq t)$ is a causal functional of the signal x_t and $F_t^y = \sigma\{y_s, 0 \leq s \leq t\} \equiv$ sub- σ -algebra generated by $y_s, 0 \leq s \leq t$, then we are interested in calculating the optimal least squares estimate of $f_t(x_{\cdot})$

$E\{f_t(x_{\cdot}) | F_t^y\}$ for $t \leq T$.

Definition 1 y_t defined in (1) and (2) is called an observation semi-martingale. Throughout, let (Ω, F, P) denote the underlying probability space.

Now $E\{f_t(x_{\cdot}) | F_t^y\} \in L(\Omega, F_t^y, P) (= \{F_t^y \text{ measurable rv's}\})$ by the definition of conditional expectation, and, therefore, any method that represents elements of $L(\Omega, F_t^y, P)$ in a simple and consistent way, say by expansion in terms of a simple class of functionals of y_{\cdot} can be applied to the optimal estimate. In this work, we have adopted multiple integrals of the form $\int_0^t \dots \int_0^{s_{r-1}} k(t, s_1, \dots, s_r) dy(s_r) \dots dy(s_1)$

as the basic objects of expansion. First, y_t is a stochastic translation of Brownian motion and through a change of measure, much Brownian theory can be carried over. Secondly, iterated integrals provide the natural concept of a polynomial in the y process and thus they give a framework for considering best quadratic, cubic, etc. suboptimal estimation procedures. Finally, when the kernel k of

$z_t = \int_0^t \int_0^{s_{r-1}} k dy_{s_r} \dots dy_{s_1}$ is separable, a construction of Brockett [2] realizes z_t recursively as the solution

to a stochastic differential equation.

Accordingly, after developing some theory of multiple integral expansions we show how $E\{f_t(x_{\cdot}) | F_t^y\}$ can be represented as a ratio of multiple integral expansions. The chief theoretical result about multiple integrals, the multiplication formula of theorem 2, is then used in conjunction with this representation to derive equations for the best suboptimal estimate of any order. The Kalman filter is derived and the quadratic filter discussed in detail as examples.

Multiple Integrals. In what follows, let $(b(t), F_t)$ denote a standard Brownian motion w.r.t. increasing family of sub- σ -algebras F_t . We assume familiarity with the stochastic integral $\int_0^t \phi_s db(s)$, where $\phi_s(u)$ is a measurable process adapted to $(F_t)_{t \geq 0}$.

Definition 2 Let $f \in L^2([0, T]^n) \equiv \{f \in L^2([0, T]^n) | f \text{ symmetric}\}$. $I_t^{(n)}(f)$, the n th order multiple (or iterated) integral up to $t \leq T$ of f , is defined recursively by

$$I_t^{(n)}(f) = \int_0^t I_s^{(n-1)}(f(s, \cdot)) db(s) \quad (3)$$

In (3), $f(s, \cdot)$ is the function of $L^2([0, T]^{n-1})$ formed by holding the first element of f fixed at s . Strictly speaking, for (3) to make sense it must be shown that $I_s^{(n-1)}(f(s, \cdot))$ has a measurable version, but this can easily be done by approximating f with separable functions. Let $(f, g) = \int_0^t \int_0^{s_{n-1}} f(s_1, \dots, s_n) g(s_1, \dots, s_n) ds_1 \dots ds_n$ denote the inner product of $L^2([0, T]^n)$. By applying standard facts about stochastic integrals, the following basic properties of the multiple integral are derived: for any n and m , $t \leq T$, and $f \in L^2([0, T]^n)$, $g \in L^2([0, T]^m)$

$$\begin{aligned} \text{a) } E\{I_t^{(n)}(f)\} &= 0 \\ \text{b) } E\{I_t^{(n)}(f) I_t^{(m)}(g)\} &= \begin{cases} 0 & \text{if } m \neq n \\ 1/n! (f, g) & \text{if } m = n \end{cases} \end{aligned} \quad (4)$$

Note also that $I_t^{(n)}(f)$ depends only on the values of $f(s_1, \dots, s_n)$ for $s_1 \leq s_2 \leq \dots \leq s_n$. (3) adopts the useful convention of allowing f to be defined in all of $[0, T]^n$ by a symmetric extension.

Multiple integrals are useful in constructing Wiener's homogeneous chaos expansion, which as an example of the general theory presented later, decomposes $L^2(F_t^y)$ into a direct sum of Hilbert space tensor products. Indeed if $H_0 \equiv \mathbb{R}$, $H_n \equiv \{I_t^{(n)}(f) | f \in L^2([0, T]^n)\}$ $n \geq 1$ a simple application of (4) a) and b) demonstrates that H_n is a Hilbert space for every n and that $H_n \perp H_m$ for $n \neq m$ [where \perp is defined in the sense of the inner product $(x, y) = Exy$]. In fact we have more:

Theorem 1 (Ito-Wiener)

$$L^2(F_t^y) = H_0 \oplus H_1 \oplus H_2 \oplus \dots$$

That is, for $\phi \in L^2(F_t^y)$ kernels $\{k_n\}_{n=0}^\infty$ exist such that

$$\phi = k_0 + \sum_{n=1}^\infty I_t^{(n)}(k_n) \quad (5)$$

Proof. See Ito [3] and Kallianpur [4].

Theorem 1 suggests the following natural question.

Suppose $f \in L^2([0, T]^n)$ and $g \in L^2([0, T]^m)$. Is it then true that $I_t^{(n)}(f) I_t^{(m)}(g) \in L^2(F_t^y)$ for $t \leq T$, and if so, what are the kernels $\{k_1\}$, as in (5), such that $I_t^{(n)}(f) I_t^{(m)}(g)$

$=k_0 + \sum_{i=1}^{\infty} I_t^{(1)}(k_i)$? Our answer, which will become a principal tool of investigation, requires some preliminary notation.

Definition 3 i) P_n will denote the projection of $L^2([0, T]^n)$ onto $L_2([0, T]^n)$

$$(P_n h)(s_1 \dots s_n) = \sum_{\pi \in S(n)} h(s_{\pi(1)} \dots s_{\pi(n)})$$

where S_n = permutation group on n letters.

ii) Let $0 \leq k \leq \min(m, n)$ $f \in L^2([0, T]^n)$, $g \in L^2([0, T]^m)$

$$(f \otimes_k (t)g)(s_1 \dots s_{m+n-2k}) = \frac{1}{k!} \int_0^t \dots \int_0^t f(r_1 \dots r_k, s_1 \dots s_{n-k}) g(r_{k+1} \dots r_{m+n-2k}) dr_k \dots dr_1$$

$$(f \otimes_k (t)g)(s_1 \dots s_{m+n-2k}) = (P_{n+n-2k} [f \otimes (t)g])(s_1 \dots s_{m+n-2k}) \quad (6)$$

To illustrate, if $n \geq m$, then, using the symmetry of f and g , $(f \otimes_m (t)g)(s_1 \dots s_{n-k}) = \frac{1}{k!} \int_0^t \dots \int_0^t f(r_1 \dots r_m, s_1 \dots s_{n-m})$

$\times g(r_{m+1} \dots r_{m+n-2k}) dr_m \dots dr_1$. It is useful to think of the functions f and g as tensors, for, in fact $L^2([0, T]^n) \cong L^2([0, T]) \otimes L^2([0, T])$ (n times). Therefore, as inspection of (6) and (7) suggests, $\otimes_k(t)$ may be interpreted as a k -fold symmetrized tensor contraction.

Theorem 2. Let $f \in L^2([0, T]^n)$, $g \in L^2([0, T]^m)$. Then $I_t^{(n)}(f) I_t^{(m)}(g) \in L^2(F_t^b)$ for $t \leq T$ and,

$$I_t^{(n)}(f) I_t^{(m)}(g) = \sum_{k=0}^{\min(m, n)} I_t^{(m+n-2k)} [\otimes_{m-k}^{m+n-2k} f \otimes_k(t)g] \quad (8)$$

Before sketching a proof, let us first demonstrate that the l.h.s. of (8) is well-defined.

Lemma 1. Let $f \in L^2([0, T]^n)$, $g \in L^2([0, T]^m)$. For $t \leq T$ $f \otimes_k(t)g \in L^2([0, T]^{m+n-2k})$.

$$\|f \otimes_k(t)g\|_{m+n-2k}^2 \leq C \|f\|_n^2 \|g\|_m^2 \quad (9)$$

where C depends on m , n and k .

Proof. Let $|S_n|$ = cardinality (S_n) and $j = m+n-2k$.

Using Cauchy-Schwarz repeatedly:

$$\|f \otimes_k(t)g\|_j \leq \frac{|S_j|}{(j!)^2} \sum_{\pi \in S(j)} \|f \otimes_k(t)g(s_{\pi(1)} \dots s_{\pi(j)})\|_j^2$$

$$\|f \otimes_k(t)g(s_{\pi(1)} \dots s_{\pi(j)})\|_j^2 = \frac{1}{(k!)^2} \int_0^t \dots \int_0^t ds_{\pi(1)} \dots ds_{\pi(j)}$$

$$\times \left[\int_0^t \dots \int_0^t f(s_1 \dots s_r, s_{\pi(1)} \dots s_{\pi(j)}) g(r_{k+1} \dots r_{m+n-2k}) dr_r \dots dr_1 \right]^2$$

$$\leq \frac{1}{(k!)^2} \|f\|_n^2 \|g\|_m^2$$

$$\text{Thus } \|f \otimes_k(t)g\|_j \leq \frac{|S_j|}{(j!)^2} \|f\|_n^2 \|g\|_m^2$$

Proof of theorem 2*: Only a sketch will be given, as details are involved and unrevealing. First, it suffices to treat the case when f and g are separable, since we can use lemma 1 to approximate general f and g by separable functions. This makes questions concerning the interchange of dt and $db(t)$ integrations easy to resolve. The case $n=1$, $m=1$ follows directly by applying Ito's differentiation rule. Indeed, Ito's rule yields in general

$$I_t^{(n)}(f) I_t^{(m)}(g) = \int_0^t I_s^{(m)}(g) I_s^{(n-1)}(f(s, \dots)) db(s) + \int_0^t I_s^{(n)}(f) I_s^{(m-1)}(g(s, \dots)) db(s) + \int_0^t I_s^{(n-1)}(f(s, \dots)) \times I_s^{(m-1)}(g(s, \dots)) ds \quad (10)$$

Using (10) we can implement the following two stage induction scheme to prove the theorem for all m and n .

*Apparently, Japanese workers have also recently proved theorem 2 by means of functional analytic techniques due to Hida (personal communication from T. Hida)

- a) The case $(m, n) = (k, 1)$ implies the case $(m, n) = (k+1, 1)$
b) The cases $(m, n) = (k-1, j)$, $(k, j-1)$ and $(k-1, j-1)$ imply the case (k, j) .

Equation (8), the multiplication formula, is actually a generalization of similar looking Hermite polynomial identity

$$h_m(x) h_n(x) = \sum_{k=0}^{\min(m, n)} \binom{m}{k} \binom{n}{k} \frac{r^{r+q-2k}}{r-k} h_{m+n-2k}(x) \quad (11)$$

where $h_n(x) = \frac{(-1)^n}{\sqrt{n!}} e^{-x^2/2} (d^n/dx^n) e^{x^2/2}$. To understand

the connection, observe that the polynomials $h_m(x)$

provide an alternate means of constructing the

decomposition of theorem 1. In fact, if $\{\phi_i\}_{i=1}^{\infty}$

is a complete orthonormal basis of $L^2([0, T])$ and

$$G_n \equiv \text{Span} \left\{ \int_0^t h_{p_1}(\phi_{i_1}(s)) db(s) \dots \int_0^t h_{p_r}(\phi_{i_r}(s)) db(s) \mid p_1 + \dots + p_r = n, i_1, \dots, i_r \text{ pairwise unequal} \right\}$$

then Ito [3] has shown that $H_n = \overline{G_n}$ ($\overline{}$ denotes closure).

(See also Kallianpur [4]). Thus a typical element

$I_t^{(n)}(f) \in H_n$ is a generalization of a Hermite polynomial.

The slight discrepancy between the factors in (11) and

(8) arises from the normalizations involved in the

definitions of h_n , $I_t^{(n)}$ and \otimes_k .

(8) has consequences that relate directly to the

theory of contractions on sums of Hilbert space tensor

products presented in a later section. The point is

that the multiplication formula can be used to study

the integrability of k th order moments of the integral

$I_t^{(m)}(f)$, and, indeed, a direct application of (8) using

lemma 1 and a recursion argument yields:

Theorem 3 Let $n \geq 1$ and $f \in L^2([0, T]^n)$. For any $k \geq 1$, there exists $M_{m,k} > 0$, independent of f , such that

$$E[I_T^{(n)}(f)]^{2k} \leq M_{m,k} \|f\|_m^{2k} \quad (12)$$

Now, Segal [5] has previously derived (12) by tensor

product operator arguments, and, in addition, proves

there exists a constant c such that $M(m, k)$ may be

replaced by k^{2ckn} . His argument thus connects (12)

to a deeper general theory.

Theorem 3 has an interesting corollary.

Theorem 4 Let $\{f_m\}_{m=1}^{\infty}$ and f be functions in $L^2([0, T]^n)$.

Then $\lim_{m \rightarrow \infty} \|f_m - f\|_n^2 = 0$ iff $\lim_{m \rightarrow \infty} E[I_T^{(n)}(f_m) - I_T^{(n)}(f)]^2 = 0$

iff $\lim_{m \rightarrow \infty} E[I_T^{(n)}(f_m) - I_T^{(n)}(f)]^{2k} = 0$ for any or all k .

Proof By (4b) $E[I_T^{(n)}(f_m) - I_T^{(n)}(f)]^2 = \frac{1}{n!} \|f_m - f\|_n^2$.

Using (12) completes the proof.

In the applications, we shall want to discuss multiple integrals not with respect to Brownian motion, but to an observation semi-martingale y_t . We again

denote these integrals by $I_t^{(n)}(f)$ without explicitly indicating the dependence on y_t , which should be

clear from the context of their use. The simplest

definition of such an integral uses a result stated in

theorem 2; namely, under condition (2) there exists

a measure P_0 on (Ω, F) such that i) y_t is Brownian

on (Ω, F, P_0) , and ii) P_0 and P are mutually absolutely

continuous.

Definition 4 For $f \in L^2([0, T]^n)$

$$I_t^{(n)}(f) = \int_0^t \dots \int_0^t f(s_1 \dots s_n) dy_{s_1} \dots dy_{s_n}$$

is a r.v.a.s. equal to the Brownian multiple integral defined on (Ω, F, P_0) .

By absolute continuity, $I_t^{(n)}(f)$ is a well-defined r.v. on (Ω, \mathcal{F}, P) . Also, as further argument will show, $I_t^{(n)}(f)$ equals the iterated integral defined directly on (Ω, \mathcal{F}, P) in the manner of definition 2, and thus the 'natural' interpretation of $I_t^{(n)}(f)$ as an iteration is preserved. It immediately follows from definition 4 that the multiplication formula holds for the observation semi-martingale case. Likewise, if $E_0(\frac{dP}{dP_0})^{2k} < \infty$ then $E[I_t^{(n)}(f)]^{2k} = E_{\frac{dP}{dP_0}}[I_t^{(n)}(f)]^{2k} \leq (E_0(\frac{dP}{dP_0})^{2k})^{1/2} (E_0[I_t^{(n)}(f)]^{2k})^{1/2}$ shows that theorem 3 extends as well.

Finally, it is important to compute the mean and variance of integrals with respect to y_t .

Lemma 2 Let $E(\int_0^T h^2(s) ds)^n < \infty$. Then for $k \leq n$ and $f \in L^2([0, T]^k)$,

$$i) E[I_t^k(f)]^2 \leq \|f\|_k^2 (M_T - \text{does not depend on } f)$$

$$ii) E[I_t^k(f)] = \int_0^t \dots \int_0^{s_{k-1}} f(s_1, \dots, s_k) E[h(s_1) \dots h(s_k)] dy_1 \dots dy_k$$

Proof The proof proceeds by induction on the order k , and the induction stops at $k=m$ because of the condition $E(\int_0^T h^2(s) ds)^n < \infty$. Details will not be presented for lack of space.

Filter expansions and applications. We will now show that the Kallianpur-Striebel formula, (13), for the optimal estimate can be developed into a representation of the estimate as a ratio of two multiple integral expansions. This technique bears comparison to the work of Eterno [6], who derived similar expressions in an effort to approximate the conditional distribution of the signal given the observation process. Here we focus on the use of the expansion to derive equations for suboptimal filters.

Recall the filtering problem 1)-2). Denote $h(x(s))$ by $h(s)$,

$f_t(x(\cdot))$ by $f(t)$, and $E\{f(t)|F_t^y\}$ by $\hat{f}(t)$, and define

$$L(t) = \exp[\int_0^t h(s) dy(s) - 1/2 \int_0^t h^2(s) ds].$$

Also, define a new measure P_0 on (Ω, \mathcal{F}) by

$$\frac{dP_0}{dP} = \exp[-\int_0^T h(s) dy(s) - 1/2 \int_0^T h^2(s) ds].$$

Theorem 5 Under the hypothesis of (2)

- i) P_0 is a probability measure and P and P_0 are mutually absolutely continuous with $\frac{dP}{dP_0} = L(T)$.
- ii) Under P_0 $y(t)$ is a Brownian motion independent of $x(t)$.
- iii) $x(t)$ has the same distribution under P_0 as under P .
- iv) (Kallianpur, Striebel)

$$E\{f(t)|F_t^y\} = E_0\{f(t) \frac{dP}{dP_0} | F_t^y\} / E_0\{\frac{dP}{dP_0} | F_t^y\} = E_0\{f(t)L(t) | F_t^y\} / E_0\{L(t) | F_t^y\}. \quad (13)$$

Proof See Wong [7],

$$\text{Let } L_{r-1}(t) = \int_0^t \dots \int_0^{s_{r-1}} h(s_1) \dots h(s_r) L(s_r) dy(s_r) \dots dy(s_1).$$

Theorem 6 a) Partial expansion

$$\text{Suppose } E[\int_0^T h^2(s) ds]^r < \infty \text{ and } E[f^2(s) \int_0^T h^2(s) ds]^r < \infty. \text{ Then } \hat{f}(t) = \frac{E(f(t) + \sum_{j=1}^r I_t^{(j)}(k_j) + E_0\{f(t)L(t) | F_t^y\})}{1 + \sum_{j=1}^r I_t^{(j)}(l_j) + E_0\{L(t) | F_t^y\}}, \quad (14)$$

where $k_j(t, s_1, \dots, s_j) = E\{f(t)h(s_1) \dots h(s_j)\}$ and

$$l_j(s_1, \dots, s_j) = E[h(s_1) \dots h(s_j)]$$

b) Full expansion. If $E[\exp[\int_0^T h^2(s) ds]] < \infty$ and

$E f^2(t) \exp[\int_0^T h^2(s) ds] < \infty$, then

$$f(t) = \frac{E(f(t) + \sum_{j=1}^{\infty} I_t^{(j)}(k_j))}{1 + \sum_{j=1}^{\infty} I_t^{(j)}(l_j)}, \quad (15)$$

where k_j and l_j are as above and the infinite series both converge in the $L^1(P)$ norm.

Remarks 1. The kernels k_j and l_j depend only on the a priori distribution of $x(t)$.

2. The condition $E[\exp[\int_0^T h^2(s) ds]] < \infty$ in (6) places strong restrictions on the growth of the moments of $\int_0^T h^2(s) ds$. Moreover

$$\begin{aligned} E_0(\frac{dP}{dP_0})^2 &= E_0 E_0\{\exp[2\int_0^T h(s) dy(s) - \int_0^T h^2(s) ds] | F_t^x\} \\ &= E_0 \exp[-\int_0^T h^2(s) ds] E_0\{\exp[2\int_0^T h(s) dy(s) | F_t^x\} \\ &= E_0 \exp[\int_0^T h^2(s) ds] \end{aligned}$$

since $\int_0^T h(s) dy(s)$ conditioned on F_t^x is normal with variance $\int_0^T h^2(s) ds$.

3. Theorem 6 can be generalized without difficulty to vector valued processes. Proof of theorem 6: For lack of space we give only an heuristic sketch. The principal idea comes from observing that, by using Ito's differentiation rule

$$\begin{aligned} dL(t) &= h(t)L(t)dy(t). \text{ In other symbols,} \\ L(t) &= 1 + \int_0^t h(s)L(s)dy(s) \end{aligned} \quad (16)$$

By iterating (16):

$$L(t) = 1 + \int_0^t h(s) dy(s) + \int_0^t \int_0^s h(s)h(r)L(r)dy(r)dy(s).$$

Continuing such iteration ad infinitum we derive the formal expression

$$L(t) = 1 + \sum_{j=1}^{\infty} \int_0^t \dots \int_0^{s_{j-1}} h(s_1) \dots h(s_j) dy(s_j) \dots dy(s_1). \quad (17)$$

Now substitute (17) into the term $E_0[L(t)|F_t^y]$. We get:

$$\begin{aligned} E_0[L(t)|F_t^y] &= 1 + \sum_{j=1}^{\infty} E_0\{\int_0^t \dots \int_0^{s_{j-1}} h(s_1) \dots h(s_j) dy(s_j) \dots dy(s_1) | F_t^y\} \\ &= 1 + \sum_{j=1}^{\infty} \int_0^t \dots \int_0^{s_{j-1}} E_0\{h(s_1) \dots h(s_j) | F_t^y\} dy(s_j) \dots dy(s_1) \\ &= 1 + \sum_{j=1}^{\infty} \int_0^t \dots \int_0^{s_{j-1}} E_0\{h(s_1) \dots h(s_j)\} dy(s_j) \dots dy(s_1) \\ &= 1 + \sum_{j=1}^{\infty} I_t^{(j)}(l_j). \end{aligned} \quad (18)$$

The second equality uses a stochastic 'Fubini' theorem found, for example, in Liptser and Shiriyayev [8]; for the process $\phi(s)$ adapted to the Brownian motion

(b_t, F_t) and satisfying $E[\int_0^T \phi^2(s) ds] < \infty$

$$E[\int_0^T \phi(s) db(s) | F_t^b] = \int_0^T E[\phi(s) | F_s^b] db(s).$$

The third equality follows from Theorem 5 ii) and iii), and the fourth equality by definition. By a similar calculation,

$$E_0\{f(t)L(t) | F_t^y\} = 1 + \sum_{j=1}^{\infty} I_t^{(j)}(k_j). \quad (19)$$

Now (18) and (19) can be substituted into

$$f(t) = E_0\{f(t)L(t) | F_t^y\} / E_0\{L(t) | F_t^y\}$$

to formally derive theorem 6, b). The partial expansion if proved by carrying out the iteration procedure of (16) only a finite number of times. The various hypotheses in theorem 6 merely guarantee that the steps in each calculation are valid.

3. Applications

The explicit formulas (14) and (15) can be applied to the design of suboptimal filters in various ways. For example, one naive approach would be to truncate the numerator and denominator of the ratio at finite orders and use the result as an approximate filter. As noted in the remarks after theorem 6, the kernels of the expansions do not involve the observations $y(\cdot)$ and so can be computed off line. Theoretically then, it is possible to construct the truncated filter. This design, however, is difficult to analyze and assess; a more interesting use of theorem 6 involves finding estimates that are multiple integral expansions of finite order.

Definition 5. a) An expression

$$c(t) = c_0(t) + \sum_{n=1}^r I_t^{(n)}(a_n(t))$$

with $c_n(t) \in L^2([0, T]^n)$ is called an r th order expansion of $y(\cdot)$.

b) An r th order expression $a(t)$, satisfying

$$E[f(t) - a(t)]^2 \leq E[f(t) - c(t)]^2$$

for any other r th order $c(t)$, is called the best r th order estimate of $f(t)$.

The best r th estimate will be denoted by $\hat{f}(t)$ (with r understood), and its kernels by a_0, a_1, \dots, a_r .

Given an order r , how can we find $\hat{f}(t)$, that is how can we determine a_0, a_1, \dots, a_r ? As it turns out, we can apply the multiplication formula to the filter expansion to write integral equations for the kernels a_n . Begin by considering the product $f(t)E_0[L(t)|F_t^Y]$ of the estimate with the denominator of (13). If

$$E[\int_0^T h^2(s) ds]^{2r} < \infty$$

then the expansion of $E_0[L(t)|F_t^Y]$ in (14) applies, and

$$f(t)E_0[L(t)|F_t^Y] = [a_0(t) + \sum_{n=1}^r I_t^{(n)}(a_n(t))] [1 + \sum_{n=1}^{2r} I_t^{(n)}(l_n(t)) + E_0\{L_{2r}(t)|F_t^Y\}] = g_0(t) + \sum_{n=1}^{3r} I_t^{(n)}(g_n(t)) + f(t)E_0\{L_{2r}(t)|F_t^Y\}. \quad (20)$$

The g_n , $n=1, \dots, 3r$ are calculated from $a_n(t)$ and $l_n(t)$ by use of the multiplication formula.

Theorem 7. Suppose $E[\int_0^T h^2(s) ds]^{2r} < \infty$, $Ef^2(t) < \infty$

$$\text{and } Ef^2(t) [\int_0^T h^2(s) ds]^{2r} < \infty.$$

Then $\hat{f}(t) = a_0(t) + \sum_{n=1}^r I_t^{(n)}(a_n(t))$ is the best r th order estimate iff

$$g_0(t) = Ef(t) \quad (21)$$

$$g_n(t, s_1, \dots, s_n) = E[f(t)h(s_1) \dots h(s_n)] = k_n, \quad 1 \leq n \leq r.$$

Proof: We must show that

$$E[f(t) - \hat{f}(t)]^2 \leq E[f(t) - c(t)]^2 \quad (22)$$

for all n th order expansions $c(t)$ iff (21) holds. Recall that $\hat{f}(t)$ may be interpreted as the projection of $f(t)$ onto $L^2(\Omega, F_t^Y, P)$. Thus the projection theorem implies

$$E[f(t) - \hat{f}(t)]^2 = E[f(t) - \hat{f}(t)]^2 + E[\hat{f}(t) - \tilde{f}(t)]^2 + 2E[f(t) - \hat{f}(t)][\hat{f}(t) - \tilde{f}(t)] = E[f(t) - \hat{f}(t)]^2 + E[\hat{f}(t) - \tilde{f}(t)]^2$$

Applying this calculation to the r.h.s. of (22) also, (22) holds iff

$$E[\hat{f}(t) - \tilde{f}(t)]^2 \leq E[\hat{f}(t) - c(t)]^2 \quad (23)$$

for all $c(t)$. But according to lemma 2, the set of r th order expansions in $y(\cdot)$ is a Hilbert space, and thus, applying the projection theorem again, (23) holds iff

$$E[\hat{f}(t) - \tilde{f}(t)]c(t) = 0 \quad (24)$$

for all r th order expansions $c(t)$. Now substitute the expression (13) for $\tilde{f}(t)$ into (24):

$$E[\hat{f}(t) - \tilde{f}(t)]c(t) = E\left[\frac{E_0\{f(t)L(t)|F_t^Y\} - \tilde{f}(t)E_0\{L(t)|F_t^Y\}}{E_0\{L(t)|F_t^Y\}}\right]c(t) \\ = E\left[\frac{dP}{dP_0}\bigg|F_t^Y\right] [E_0\{f(t)L(t)|F_t^Y\} - \tilde{f}(t)E_0\{L(t)|F_t^Y\}]c(t) \\ = E_0\{E_0\{f(t)L(t)|F_t^Y\} - \tilde{f}(t)E_0\{L(t)|F_t^Y\}|c(t)\} \quad (25)$$

The second equality in (25) uses the identities

$$E\left[\frac{dP}{dP_0}\bigg|F_t^Y\right] = [E_0\left\{\frac{dP}{dP_0}\bigg|F_t^Y\right\}]^{-1} \\ = [E_0\{L(t)|F_t^Y\}]^{-1},$$

which are easily demonstrated. Now under P_0 , $y(\cdot)$ is a Brownian motion and integrals of different orders are orthogonal. Thus, using (20) and

$$E_0\{\hat{f}(t)L(t)|F_t^Y\} = Ef(t) + \sum_{n=1}^r I_t^{(n)}(k_n) + E_0\{f(t)L_r(t)|F_t^Y\}$$

in (25),

$$E[\hat{f}(t) - f(t)]c(t) = E_0\{Ef(t) - g_0 + \sum_{n=1}^r I_t^{(n)}(k_n - g_n)\}c(t) \\ + E_0\{f(t)c(t)E_0\{L_{2r}(t)|F_t^Y\}\} \\ + E_0\{c(t)E_0\{f(t)L_r(t)|F_t^Y\}\} \quad (26)$$

An application of lemma 2 show that the second and third terms of the r.h.s. of (26) are zero for all $c(t)$. Clearly, the first term can be zero for all n th order $c(t)$ iff $k_n = g_n$ for $0 \leq n \leq r$, and this completes the proof.

The equations (26) are actually integral equations for the kernels $a_n(t)$ of the best r th order estimate, since the $g_n(t)$, $0 \leq n \leq r$, are found from $a_n(t)$, $0 \leq n \leq r$, by the formula (8). To illustrate, if $r=1$, $l_1(s) = Eh(s)$ and

$$Ef(t) = g_0(t) = a_0(t) + \int_0^t a_1(t, u) Eh(u) du$$

$$Ef(t)h(s) = g_1(t, s)$$

$$= a_0(t) Eh(s) + a_1(t, s).$$

Solving for $a_0(t)$,

$$a_1(t, s) + \int_0^t a_1(t, u) \text{cov}[h(s), h(u)] du = \text{cov}[f(t), h(s)].$$

This is the familiar Wiener-Hopf equation for the best linear estimate. In the best quadratic ($r=2$) case, the equations become more complicated. They are, assuming $Eh(s)=0$, $Ef(t)=0$ for simplicity,

$$a_0(t) = -\int_0^t \int_0^{u_1} a_2(t, u_1, u_2) Eh(u_1)h(u_2) ds_2 ds_1 \quad (27a)$$

$$a_1(t, s) = Ef(t)h(s) - \int_0^t a_1(t, u) Eh(s)h(u) du \\ - \int_0^t \int_0^{u_1} a_2(t, u_1, u_2) \text{cov}[h(s), h(u_1)h(u_2)] du_2 du_1 \quad (27b)$$

$$a_2(t, s_1, s_2) = \text{cov}[f(t), h(s_1), h(s_2)] \\ - \int_0^t a_1(t, u) \text{cov}[h(s_1), h(s_2), h(u)] du \\ - \int_0^t [a_2(t, s_1, u) Eh(s_2)h(u) + a_2(t, s_2, u) Eh(s_1)h(u)] du \\ - \int_0^t \int_0^{u_1} a_2(t, u_1, u_2) \text{cov}[h(s_1), h(s_2), h(u_1)h(u_2)] du_2 du_1 \quad (27c)$$

[In (27), $\text{cov}[X_1, \dots, X_n] = E(X_1 - EX_1) \dots (X_n - EX_n)$.]
 (27) shows how the kernels of different orders are dependent on one another. Though not a standard integral equation, (27) may be reduced, by using the solution of a related linear estimation problem, to a single integral equation for a_2 . For fixed t this equation is of Fredholm type for $a_2(t, \cdot, \cdot)$ and can be solved by standard methods. We shall not go into this theory here.

The multiplication formula can also be used to derive the Kalman filter. Consider the simple case

$$dx(t) = db(t) \quad x(0) = 0$$

$$dy(t) = x(t)dt + dw(t)$$

where $b(\cdot)$ and $w(\cdot)$ are independent Brownian motions. Then we can show that the optimal filter is

$$\hat{x}(t) = \int_0^t a(t, s) dy(s)$$

where

$$a_1(t, s) + \int_0^t a_1(t, u) E b(u) b(s) = E b(t) b(s). \quad (28)$$

The proof is simply to show that $a(t, s)$ can be chosen so that

$$\hat{x}(t) = \int_0^t a(t, s) dy(s) = E f(t) + \sum_{j=1}^{\infty} I_t^{(j)}(k_j) / 1 + \sum_{j=1}^{\infty} I_t^{(j)}(1_j) \quad (29)$$

or

$$\int_0^t a(t, s) dy(s) [1 + \sum_{j=1}^{\infty} I_t^{(j)}(1_j)] = E f(t) + \sum_{j=1}^{\infty} I_t^{(j)}(k_j). \quad (30)$$

By expanding the l.h.s of (29) using (8), and equating kernels of different orders we derive the infinite set of equations.

$$j a(t, \cdot) \theta_0(t) 1_{j-1} + a_1(t, \cdot) \theta_1(t) 1_{j+1} = k_j. \quad (31)$$

It can now be shown that if (31) is satisfied for $j=1$, it is satisfied for all $j \geq 1$, a result following from the identity for Gaussian random variables:

$$1_j(s_1 \dots s_j) = E b(s_1) \dots b(s_j) \\ = \sum_{j=2}^n \text{cov}(b(s_1) b(s_2)) E [b(s_2) \dots b(s_{j-1}) b(s_{j+1}) \dots b(s_j)]$$

(see Miller [9]). This derivation is somewhat formal because the condition for the full expansion in (29)

to hold is that $E[\exp \int_0^t b^2(s) ds] < \infty$, which is valid only for small t .

4. Relationship to Second Quantization (After Segal and Nelson).

Let H be a real Hilbert space and let $F: H \rightarrow RV(\Omega, A, \mu)$ be the unit Gaussian determined random field. If f_1, \dots, f_n are orthonormal in H and ϕ is a Bounded Baire function on R^n , then

$$\int \phi(F(f_1), \dots, F(f_n)) = \frac{1}{(2\pi)^n} \int_{R^n} \phi(x) e^{-\|x\|^2/2} dx$$

For concreteness (Ω, A, μ) may be chosen to be countably infinite copies of $(R, B(R), (2\pi)^{-1/2} e^{-x^2/2} dx)$.

If E denotes expectation on (Ω, A, μ) then

$$E(F(f_1) \dots F(f_{2n+1})) = 0 \quad (32)$$

$$E(F(f_1) \dots F(f_{2n})) = \sum [f_{i_1}, f_{j_1}] \dots [f_{i_n}, f_{j_n}] \quad (33)$$

where the sum is over all pairings of $1, \dots, 2n$, i.e.

$i_1 < \dots < i_n; j_1 < \dots < j_n$, and

$(i_1, j_1, \dots, i_n, j_n)$ is a permutation $1, \dots, 2n$.

$L^p(\Omega, A, \mu)$ is denoted by $L^p(H)$ and $\Gamma(H)$ denotes

$L^2(H)$. $\Gamma(H)_{\leq n}$ be the closed linear span in $\Gamma(H)$ of all elements of the form $F(f_1) \dots F(f_n)$ with $m \leq n$ and let $\Gamma(H)_n$ denote the orthogonal complement of

$\Gamma(H)_{\leq n-1}$ in $\Gamma(H)_{\leq n}$. For f_1, \dots, f_n in H we define the Wick polynomial:

$$:F(f_1) \dots F(f_n):$$

to be the orthogonal projection of $F(f_1) \dots F(f_n)$ into $\Gamma(H)_n$. In the special case, where H is one dimensional and hence $\Gamma(H) = L^2(R, B(R), (2\pi)^{-1/2} e^{-x^2/2} dx)$, $\Gamma(H)_n$ is the one dimensional subspace spanned by the n th Hermite polynomial and $:x^n:$ is the n th Hermite polynomial normalization so that the leading coefficient is 1.

We have the formula

$$[:F(f_1) \dots F(f_n):, :F(g_1) \dots F(g_n):] \\ = \sum_{\pi} [f_{\pi(1)}, g_1] \dots [f_{\pi(n)}, g_n]. \quad (34)$$

where the sum is over all permutations π of $1, \dots, n$. If all the f 's and g 's are equal, we get

$$[:F(f)^n, :F(f)^n:] = \frac{1}{2} \int_{-\infty}^{\infty} (:x^n:)^2 e^{-x^2/2} dx = n!. \quad (35)$$

Let H_1 be the complexification of H (and let H_n denote the n -fold Hilbert space) symmetric tensor product of H_1 with itself. On H_2 we define the inner product such that

$$[\text{Sym}(f_1 \otimes \dots \otimes f_n), \text{Sym}(g_1 \otimes \dots \otimes g_n)] = \sum_{\pi} [f_{\pi(1)}, g_1] \dots [f_{\pi(n)}, g_n] \quad (36)$$

where

$$\text{Sym}(f_1 \otimes \dots \otimes f_n) = \frac{1}{n!} \sum_{\pi} f_{\pi(1)} \otimes \dots \otimes f_{\pi(n)}. \quad (37)$$

From (34) and (36), that the mapping

$$:F(f_1) \dots F(f_n): \rightarrow \text{Sym}(f_1 \otimes \dots \otimes f_n)$$

extends uniquely to the unitary operator from $\Gamma(H)_n$ onto H_n . We use this mapping to identify $\Gamma(H)_n$ and H_n .

Analogous to the fact that the one-dimensional Hermite polynomials span $L^2(R, B(R), (2\pi)^{-1/2} e^{-x^2/2} dx)$, Segal proved

$$\Gamma(H) = \bigoplus_{n=0}^{\infty} H_n. \quad (38)$$

for arbitrary real Hilbert space H . $\Gamma(H)$ is Fock Space.

If the random field $F(f) = f dB$, where

$f \in L^2(R) = H$ and B is the standard Wiener process, the elements of $\Gamma(H)_n$ are multiple Wiener integrals (in the sense of Ito).

The space $\Gamma(H)$ is intrinsically attached to the structure of H as a real Hilbert space. Thus if $U: H \rightarrow K$ is an orthogonal mapping of one real Hilbert space into another, it induces a unitary mapping $\Gamma(U): \Gamma(H) \rightarrow \Gamma(K)$, where on H_n , $\Gamma(U) = U \otimes \dots \otimes U$. Similarly if $I: H \rightarrow K$ is an isometric embedding then it induces an isometric embedding $\Gamma(I): \Gamma(H) \rightarrow \Gamma(K)$ and similarly for an orthogonal projection $E: H \rightarrow K$. If $A: H \rightarrow K$ is a contraction then $\Gamma(A): \Gamma(H) \rightarrow \Gamma(K)$ is defined to be the direct sum to $\Gamma(A)_n: H_n \rightarrow K_n$, where $\Gamma(A)_n = A \otimes \dots \otimes A$. Now any contraction $A: H \rightarrow K$ can be decomposed as

$$\begin{array}{ccccc} H & \xrightarrow{I} & H \otimes K & \xrightarrow{U} & K \otimes H \\ & \searrow A & & \nearrow E & \\ & & K & & \end{array}$$

where I , U and E are as above.

Hence $\Gamma(A) = \Gamma(E) \Gamma(U) \Gamma(I)$. Now $\Gamma(A)$ is doubly Markovian in the sense that

$$\alpha \geq 0 \rightarrow \Gamma(A) \alpha \geq 0 \\ \Gamma(A) 1 = 1 \\ E \Gamma(A) \alpha = E \alpha. \quad (39)$$

Any doubly Markovian operation is a contraction from L^p to L^p .

It turns out that $\Gamma(A)$ has stronger contractive

properties and the precise statement of this is an important theorem of Nelson. Before we discuss this result it is useful to recall that conditional

expectations on $L^2(\Omega, \mathcal{A}, u)$ can be characterised as linear positivity preserving operators which are idempotent, of norm ≤ 1 and preserve constants. We also know that for $p \in [1, \infty]$, $p \neq 2$, all linear operators T on $L^p(\Omega, \mathcal{A}, u)$, which are idempotent, contracting and such that $T1=1$ is necessarily a conditional expectation.

Theorem 3.1 (Nelson Hypercontractivity Theorem).

Let $A: H \rightarrow K$ be a contraction. Then $\Gamma(A)$ is a

contraction from $L^q(H) \rightarrow L^p(K)$ for $1 \leq q \leq p \leq \infty$ provided that

$$\|A\| \leq \left(\frac{q-1}{p-1}\right)^{1/2} \quad (40)$$

If (40) does not hold then $\Gamma(A)$ is not a bounded operator from $L^q(H) \rightarrow L^p(K)$.

- [1] M. Fujisaki, G. Kallianpur and H. Kunita, "Stochastic Differential Equations for the Nonlinear Filtering Problem," Osaka J. Math., Vol. 9, 1972, pp. 19-40.
- [2] R.W. Brockett, "Volterra Series and Geometric Control Theory," Automatica, 12 (1976), pp. 167-176.
- [3] K. Ito, "Multiple Wiener Integrals," J. Math. Soc. Japan, 13 (1951) pp. 157-169.
- [4] G. Kallianpur, "The Role of Reproducing Kernel Hilbert Spaces in the Study of Gaussian Processes," Advances in Probability, Vol. 2, ed. P. Ney, Marcel Dekker, N.Y., 1970.
- [5] Segal, "Construction of Non-Linear Local Quantum Processes: I," Ann. Math., 92 (1970), pp. 462-481.
- [6] J.S. Eterno, Nonlinear Estimation Theory and Phase-Lock Loops, Ph.D. thesis, MIT, 1976.
- [7] E. Wong, Stochastic Processes in Information and Dynamical Systems, McGraw Hill, New York, 1971.
- [8] R.S. Liptser and A.N. Shiriyayen, Statistics of Random Processes, Vol. 1, Springer-Verlag, New York, 1979.
- [9] K.S. Miller, Multidimensional Gaussian Distributions, John Wiley, New York 1965.

Acknowledgment

This research has been supported by the Air Force Office of Scientific Research under Grant No. AFOSR-77-32816